Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística

Inference for a Skew Extension of the Gubbs Measurement Error Model

Lourdes C. Montenegro, Heleno Bolfarine e Víctor H. Lachos

Relatório Técnico

RTP-03/2006

Série Pesquisa

INFERENCE FOR A SKEW EXTENSION OF THE GRUBBS MEASUREMENT ERROR MODEL

Lourdes C. Montenegro[1]
Departamento de Estatística - ICEX -UFMG
Heleno Bolfarine[2]
Departamento de Estatística - IME - USP
V.H. Lachos[3]
Departamento de Estatística - IMECC - UNICAMP

Abstract

In this paper we discuss inferential aspects for the Grubbs model when the unknown quantity x (latent variable) follows a skew-normal distribution, extending early results given in Arellano-Valle et al. (2005b). Maximum likelihood parameter estimates are computed via the EM-algorithm. Wald and likelihood ratio type statistics are used for hypothesis testing and we explain the apparent failure of the Wald statistics in detecting skewness via the profile likelihood function. The results and methods developed in this paper are illustrated with a numerical example.

Key Words: skew-normal distribution; EM-algorithm; skewness; Grubbs model; profile likelihood.

1. Introduction

During the last decade, there has been growing interest for models that provide flexibility in capturing a broad range of non-normal behavior and hence, represent features of the data as adequately as possible and to reduce unrealistic assumptions. Motivation originated from data sets presenting clear indication of skewness and multi-modality (not following the symmetric normal law) in areas, such as, engineering, medicine, psychology and agriculture.

Although the normality assumption (or symmetry) is adequate in many situations, it is not appropriate when the data present non-normal behavior such as asymmetry. This is the case with the data set studied in Barnett (1969) which seems to require data transformation in order to be better approximated by the normal distribution. Azzalini

[1]e-mail: lourdes@est.ufmg.br
[2]e-mail: hbolfar@ime.usp.br
[3]e-mail: hlachos@ime.unicamp.br

and Capitanio (1999) list several reasons to avoid this method (variable transformation) if a more suitable theoretical model can be found: (i) it typically will not provide very useful information on an underlying data generation mechanism, (ii) the transformations are usually on each component separately, and achievement of joint normality is only hoped for; (iii) the transformed variables are more difficult to deal with regarding interpretation, especially when each variable is transformed using a different function; (iv) multivariate homoscedasticity often requires a different transformation than the one to get normality; and (v) a model for one set of data may often not be found applicable to subsequent sets. Thus, in order to accommodating such departures, we pursue a goal toward flexibility and tractability, adopting a general class of models which comprises the normal one as a proper element.

The skew-normal family was introduced by Azzalini (1985) in a univariate context, whereas Azzalini and Dalla-Valle (1996) introduced the multivariate version. In this family, a shape parameter regulates the skewness of the distribution, allowing for a continuous transition from non-normality to normality. Since the pioneer work of Azzalini (1985) we have observed an ever growing interest in the skew-normal distribution and applications to many different models have flourished in the literature (see Arellano Valle et al., 2005a, Arellano Valle et al., 2005b; Bazan et al., 2005).

In our work we concentrate on a special case of the so-called fundamental skew-normal distribution of Arellano-Valle and Genton (2004). Specifically, here we say that a k-dimensional random vector $\mathbf{y}$ has a multivariate skew-normal distribution with location parameter $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Psi}$ (positive definite), and skewness parameter vector $\boldsymbol{\lambda}$, which will be denoted by $SN_k(\boldsymbol{\mu}, \boldsymbol{\Psi}, \boldsymbol{\lambda})$, if its probability density function is given by

$$f(\mathbf{y}) = 2\phi_k(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Psi})\Phi_1(\boldsymbol{\lambda}^\top \boldsymbol{\Psi}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})), \quad \mathbf{y} \in \mathbb{R}^k, \tag{1}$$

where $\phi_k(.|\boldsymbol{\mu}, \boldsymbol{\Psi})$ stands for the density function (pdf) of the k-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariate matrix $\boldsymbol{\Psi}$, $\boldsymbol{\Phi}_1(.)$ represents the cumulative distribution function (cdf) of the standard normal distribution, and $\boldsymbol{\Psi}^{-1/2}$ satisfies $\boldsymbol{\Psi}^{-1/2}\boldsymbol{\Psi}^{-1/2} = \boldsymbol{\Psi}^{-1}$. When $\boldsymbol{\lambda} = \mathbf{0}$, we have that $\mathbf{y} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Psi})$. The stochastic representation of a skew-normal random variable, which can be used to simulate random realizations from $\mathbf{y}$, is given by

$$\mathbf{y} \overset{d}{=} \boldsymbol{\mu} + \boldsymbol{\Psi}^{1/2}(\boldsymbol{\delta}|T_0| + (\mathbf{I}_k - \boldsymbol{\delta}\boldsymbol{\delta}^\top)^{1/2}\mathbf{T}_1), \quad \text{with} \quad \boldsymbol{\delta} = \frac{\boldsymbol{\lambda}}{\sqrt{1 + \boldsymbol{\lambda}^\top \boldsymbol{\lambda}}}, \tag{2}$$

where $T_0 \sim N_1(0, 1)$ and $\mathbf{T}_1 \sim N_k(\mathbf{0}, \mathbf{I}_k)$ are independent, and " $\overset{d}{=}$ " meaning "distributed as". For more details on this approach, see Arellano-Valle and Genton (2005) and Arellano-Valle et al. (2005a). Note in (1) that when $k = 1$ we obtain the univariate skew-normal distribution introduced by Azzalini (1985) and (2) is reduced to the stochastic representation obtained in Henze (1986).

The Grubbs model is typically used in method comparison studies to assess the relative agreement between two or more analytical methods (or instruments) that measure the same quantity of interest. The primary objective is to see wether the two or more methods produce the same error in measurement. In this paper we extend the usual normal structural Grubbs model to a class of skew-normal Grubbs model (henceforth abbreviated as SN-GM), meaning that the true covariate (quantity of interest) is distributed according to the skew-normal distribution.

The paper is organized as follows. Section 2 covers model formulation as well as some inferential results. In Section 3, we present the EM-algorithm for parameter estimation. In Section 4 we propose testing statistics for some hypothesis of interest. Finally, an illustrative example previously analyzed under normal Grubbs model (henceforth abbreviated as N-GM) is reanalyzed in Section 5.

2. The model

Suppose that we have at our disposal $p \geq 2$ instruments for measuring a characteristic of interest x in a group of n experimental units. Let x_i, the unobserved (true) covariate value corresponding to unit i and y_{ij} the measured value obtained with the instrument j in unit i, $i = 1, \ldots, n$ and $j = 1, \ldots, p$. Relating these variables we consider the model (see Grubbs, 1948, 1973, 1983),

$$\mathbf{y}_i = \mathbf{a} + \mathbf{1}_p x_i + \boldsymbol{\epsilon}_i, \tag{3}$$

$i = 1, \ldots, n$, where $\mathbf{a} = (0, \boldsymbol{\alpha}^\top)^\top = (0, \alpha_2, \ldots, \alpha_p)^\top$ and $\mathbf{1}_p = (1, \ldots, 1)^\top$ are $p \times 1$ vectors; $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})^\top$ and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{ip})^\top$ (the error vector) are $p \times 1$ random vectors. In this case, we are supposing that the first instrument is a reference one that will be compared to the remaining $p - 1$ instruments. We assume that $(x_i, \boldsymbol{\epsilon}_i^\top)^\top$ follows a $(p + 1)$-variate skew-normal distribution, that is,

$$\begin{pmatrix} x_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} \stackrel{\text{ind}}{\sim} SN_{p+1}\left(\begin{pmatrix} \mu_x \\ \mathbf{0} \end{pmatrix}, D(\phi_x, \boldsymbol{\phi}_i), \begin{pmatrix} \lambda_x \\ \mathbf{0} \end{pmatrix} \right), \tag{4}$$

$i = 1, \ldots, n$, where $D(\phi_x, \boldsymbol{\phi}_i)$ denoting a diagonal matrix with elements ϕ_x and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^\top$. According to Arellano-Valle and Genton (2004), this formulation implies that

$$\boldsymbol{\epsilon}_i \stackrel{\text{iid}}{\sim} N_p(0, D(\boldsymbol{\phi})) \text{ and } x_i \stackrel{\text{iid}}{\sim} SN_1(\mu_x, \phi_x, \lambda_x), \tag{5}$$

all independent, $i = 1, \ldots, n$. This model is considering, for instance, in the case of the data set in Barnett (1969), that the distribution of the vital capacity of the human lung is not symmetrically distributed in the population and since $\boldsymbol{\epsilon}_i$ is related to model error, is expected to be normally distributed (and with mean zero). Moreover, by supposing that the true unobserved covariate (x) follows a skew-normal distribution, our model takes a step beyond a normal structural model.

Classical inference on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\phi}^\top, \mu_x, \phi_x, \lambda_x)^\top$ in this type of model is based on the marginal distribution for the response $\mathbf{y}_i$ which, as shown next, has marginally, a multivariate skew-normal distribution.

Proposition 1. Under the Grubbs model defined in (3)-(5) it follows that the marginal distribution of $\mathbf{y}_i$ is given by

$$f(\mathbf{y}_i|\boldsymbol{\theta}) = 2\phi_p(\mathbf{y}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi_1(\bar{\boldsymbol{\lambda}}_x^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu})), \tag{6}$$

$i = 1, \ldots, n$, i.e., $\mathbf{y}_i \overset{\text{iid}}{\sim} SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \bar{\boldsymbol{\lambda}}_x)$, with $\boldsymbol{\mu} = \mathbf{a} + \mathbf{1}_p\mu_x$, $\boldsymbol{\Sigma} = D(\boldsymbol{\phi}) + \phi_x \mathbf{1}_p \mathbf{1}_p^\top$, $\bar{\boldsymbol{\lambda}}_x = \dfrac{\lambda_x \phi_x \boldsymbol{\Sigma}^{-1/2}\mathbf{1}_p}{\sqrt{\phi_x + \lambda_x^2 \Lambda_x}}$, where $\Lambda_x = ({\phi_x}^{-1} + \mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi})\mathbf{1}_p)^{-1}$.

Proof. The proof is direct by using Lemmas 1 and 2 from Arellano-Valle et al. (2005a).

Thus, the log-likelihood function for $\boldsymbol{\theta}$ given the observed sample $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$ is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}), \tag{7}$$

where

$$\ell_i(\boldsymbol{\theta}) = \log(2) - \frac{p}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}d_i + \log(K_i), \tag{8}$$

$i = 1, \ldots, n$, with $d_i = (\mathbf{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})$, $K_i = \Phi_1(\bar{\boldsymbol{\lambda}}_x^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}))$ and $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\bar{\boldsymbol{\lambda}}_x$ as in (6).

It is also true (Arellano-Valle and Genton, 2005) that

$$d_i = (\mathbf{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}) \overset{\text{iid}}{\sim} \chi_p^2, \tag{9}$$

$i = 1, \ldots, n$. Such distributional result enables checking the model adequacy in practice, as seen in Section 4. In the next section we discuss an iterative process for the parameter estimation, based on the EM algorithm.

3. Likelihood based estimation

Direct maximization of the log-likelihood function (7) through the quasi-Newton method (BFGS, implemented in Matlab, R and Ox) may sometimes pose difficulties since it involves terms like $\log \Phi(a)$, which causes computational problems for negative a ($a < -38$, for example). Further, the approach does not seem too robust with respect to starting values, that is, unless good starting values are used, the direct maximization approach will typically not converge.

The EM algorithm (Dempster, et al., 1977) is a popular iterative algorithm for ML estimation in models with incomplete data. More specifically, let $\mathbf{y}$ denote the observed data and $\mathbf{s}$ denote the missing data. The complete data $\mathbf{z} = (\mathbf{y}, \mathbf{s})$ is $\mathbf{y}$ augmented with $\mathbf{s}$, the missing data. We denote by $\ell_c(\boldsymbol{\theta}|\mathbf{y}, \mathbf{s})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the complete-data log-likelihood function and by $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m-1)})$, the expected value of the complete-data log-likelihood with respect to the unknown data $\mathbf{s}$ given the observed data $\mathbf{y}$ and the current parameter estimates. Hence, let

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m-1)}) = E[\ell_c(\boldsymbol{\theta}|\mathbf{y}, \mathbf{s})|\mathbf{y}, \boldsymbol{\theta}^{(m-1)}],$$

where $\boldsymbol{\theta}^{(m-1)}$ are the current parameters estimates used to evaluate the expectation and $\boldsymbol{\theta}$ are the new parameter values optimized to increase Q.

Each iteration of the EM algorithm involves two steps, the expectation step and the maximization step:

E-step: Compute $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m-1)})$ as a function of $\boldsymbol{\theta}$;

M-step: Find $\boldsymbol{\theta}^{(m)}$ such that $Q(\boldsymbol{\theta}^{(m)}, \boldsymbol{\theta}^{(m-1)}) = \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m-1)})$.

Each iteration of the EM algorithm increases the likelihood function $\ell(\boldsymbol{\theta})$ and the EM algorithm typically converges to a local or global maximum of the likelihood function. Hence, in order to ensure that the true maximum is identified, it is typically recommended to run the EM algorithm several times with different starting values.

Using the stochastic representation given in (2), the model defined in (3)-(4) can be written hierarchically as

$$\begin{aligned} \mathbf{y}_i \mid x_i &\overset{\text{ind}}{\sim} N_p(\mathbf{a} + \mathbf{1}_p x_i, D(\boldsymbol{\phi})), && (10) \\ x_i \mid t_i &\overset{\text{ind}}{\sim} N_1(\mu_x + \phi_x^{1/2}\delta_x t_i, \phi_x(1 - \delta_x^2)), && (11) \end{aligned}$$

and

$$t_i \overset{\text{iid}}{\sim} \mathrm{HN}_1(0, 1), \quad (12)$$

$i = 1, \ldots, n$, all independent, where $\mathrm{HN}_1(0, 1)$ denotes the standardized univariate half-normal distribution and $\delta_x = \lambda_x/(1 + \lambda_x^2)^{1/2}$. In the sequel we present the EM algorithm for the SN-GN by considering that $(\mathbf{x}, \mathbf{t})$ are missing data, i.e., using double augmentation, with $\mathbf{x} = (x_1, \ldots, x_n)^\top$ and $\mathbf{t} = (t_1, \ldots, t_n)^\top$. Hence, under the representation (10)-(12), with $\nu^2 = \phi_x(1-\delta_x^2)$ and $\varsigma = \phi_x^{1/2}\delta_x$, it follows that the complete log-likelihood function associated with $(\mathbf{y}, \mathbf{x}, \mathbf{t})$ is given by

$$\begin{aligned} \ell_c(\boldsymbol{\theta}|\mathbf{y}, \mathbf{x}, \mathbf{t}) &= constant - \frac{n}{2}\log(|D(\boldsymbol{\phi})|) - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \mathbf{a} - \mathbf{1}_p x_i)^\top D^{-1}(\boldsymbol{\phi}) \\ &\quad \times (\mathbf{y}_i - \mathbf{a} - \mathbf{1}_p x_i) - \frac{n}{2}\log(\nu^2) - \frac{1}{2\nu^2}\sum_{i=1}^{n}(x_i - \mu_x - \varsigma t_i)^2. \end{aligned}$$

Letting $\widehat{x}_i = E[x_i|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{y}_i]$, $\widehat{x_i^2} = E[x_i^2|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{y}_i]$, $\widehat{t}_i = E[T_i|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{y}_i]$, $\widehat{t_i^2} = E[T_i^2|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{y}_i]$ and $\widehat{tx_i} = E[T_i x_i|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{y}_i]$, using double conditional expectations and the moments of the truncated normal distributions (see Johnson et al., 1994, Section 10.1), we obtain

$$\begin{aligned}
\widehat{t}_i &= \widehat{\mu}_{Ti} + W_{\Phi_1}\left(\frac{\widehat{\mu}_{T_i}}{\widehat{M}_T}\right)\widehat{M}_T, &(13)\\
\widehat{t_i^2} &= \widehat{\mu}_{T_i}^2 + \widehat{M_T^2} + W_{\Phi_1}\left(\frac{\widehat{\mu}_{T_i}}{\widehat{M}_T}\right)\widehat{M}_T\widehat{\mu}_{T_i}, &(14)\\
\widehat{x}_i &= \widehat{r}_i + \widehat{s}\,\widehat{t}_i, &(15)\\
\widehat{x_i^2} &= \widehat{T_x}^2 + \widehat{r}_i^2 + 2\widehat{r}_i\,\widehat{s}\,\widehat{t}_i + \widehat{s}^2\,\widehat{t_i^2}, &(16)\\
\widehat{tx_i} &= \widehat{r}_i\,\widehat{t}_i + \widehat{s}\,\widehat{t_i^2}, &(17)
\end{aligned}$$

where $W_{\Phi_1}(u) = \phi_1(u)/\Phi_1(u)$, $u \in \mathbb{R}$,

$$\widehat{M_T^2} = \left\{1 + \widehat{\varsigma}^2\mathbf{1}_p^\top(D(\widehat{\boldsymbol{\phi}}) + \widehat{\nu}^2\mathbf{1}_p\mathbf{1}_p^\top)^{-1}\mathbf{1}_p\right\}^{-1},$$

$$\widehat{\mu}_{T_i} = \varsigma\widehat{M_T^2}\mathbf{1}_p^\top(D(\widehat{\boldsymbol{\phi}}) + \widehat{\nu}^2\mathbf{1}_p\mathbf{1}_p^\top)^{-1}(\mathbf{y}_i - \widehat{\mathbf{a}} - \mathbf{1}_p\widehat{\mu_x}),$$

$\widehat{T_x}^2 = \nu^2\left\{1 + \widehat{\nu}^2\mathbf{1}_p^\top D^{-1}(\widehat{\boldsymbol{\phi}})\mathbf{1}_p\right\}^{-1}$, $\widehat{r}_i = \widehat{\mu_x} + \widehat{T_x}^2\mathbf{1}_p^\top D^{-1}(\widehat{\boldsymbol{\phi}})(\mathbf{y}_i - \widehat{\mathbf{a}} - \mathbf{1}_p\widehat{\mu_x})$, and $\widehat{s} = \widehat{\varsigma}(1 - \widehat{T_x}^2\mathbf{1}_p^\top D^{-1}(\widehat{\boldsymbol{\phi}})\mathbf{1}_p)$, $i = 1, \ldots, n$.

Using a simple algebra we get

$$\begin{aligned}
&E[\ell_c(\boldsymbol{\theta}|\mathbf{y},\mathbf{x},\mathbf{t})|\mathbf{y},\widehat{\boldsymbol{\theta}}] = constant - \frac{n}{2}\log(|D(\boldsymbol{\phi})|)\\
&-\frac{1}{2}\sum_{i=1}^n(\mathbf{y}_i - \mathbf{a} - \mathbf{1}_p\widehat{x}_i)^\top D^{-1}(\boldsymbol{\phi})(\mathbf{y}_i - \mathbf{a} - \mathbf{1}_p\widehat{x}_i) - \frac{1}{2}\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi})\mathbf{1}_p\sum_{i=1}^n(\widehat{x_i^2} - \widehat{x}_i^2)\\
&-\frac{n}{2}\log(\nu^2) - \frac{1}{2\nu^2}\sum_{i=1}^n(\widehat{x_i^2} + \mu_x^2 + \varsigma^2\widehat{t_i^2} - 2\widehat{x}_i\mu_x - 2\varsigma\widehat{xt}_i + 2\varsigma\mu_x\widehat{t}_i). \qquad (18)
\end{aligned}$$

We have then the following EM algorithm:
E-step: Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, compute $\widehat{t}_i$, $\widehat{t_i^2}$, $\widehat{x}_i$, $\widehat{x_i^2}$ and $\widehat{tx_i}$ for $i = 1, \ldots, n$, using (13)-(17).
M-step: Update $\widehat{\boldsymbol{\theta}}$ by maximizing $E[\ell_c(\boldsymbol{\theta}|\mathbf{y},\mathbf{x},\mathbf{t})|\mathbf{y},\widehat{\boldsymbol{\theta}}]$ in (18) over $\boldsymbol{\theta}$, which leads to

$$\begin{aligned}
\widehat{\alpha}_j &= \bar{y}_j - \bar{\widehat{x}}, \quad \widehat{\phi}_1 = \frac{1}{n}\sum_{i=1}^n(y_{i1}^2 - 2\widehat{x}_i y_{i1} + \widehat{x_i^2}), \quad \widehat{\mu}_x = \frac{1}{n}\sum_{i=1}^n(\widehat{x}_i - \varsigma\widehat{t}_i),\\
\widehat{\phi}_j &= \frac{1}{n}\sum_{i=1}^n(y_{ij}^2 + \alpha_j^2 + \widehat{x_i^2} - 2\alpha_j y_{ij} - 2y_{ij}\widehat{x}_i + 2\alpha_j\widehat{x}_i),\\
\widehat{\nu}^2 &= \frac{1}{n}\sum_{i=1}^n(\widehat{x_i^2} + \mu_x^2 + \varsigma^2\widehat{t_i^2} - 2\mu_x\widehat{x}_i - 2\varsigma\widehat{tx}_i + 2\varsigma\mu_x\widehat{t}_i), \quad \widehat{\varsigma} = \frac{\sum_{i=1}^n(\widehat{tx}_i - \mu_x\widehat{t}_i)}{\sum_{i=1}^n\widehat{t_i^2}},
\end{aligned}$$

where $\bar{y}_j = \frac{1}{n}\sum_{i=1}^n y_{ij}$, $\bar{\widehat{x}} = \frac{1}{n}\sum_{i=1}^n\widehat{x}_i$ and $j = 2, \ldots, p$. The skewness and scale parameters of the latent variable (x), can be estimated by noting that $\varsigma/\nu = \lambda_x$, and $\phi_x = \varsigma^2 + \nu^2$.

Starting values for $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}^0}$, can be taken from the estimates in the model normal, with values to $\widehat{\lambda_x^0} > 0$ ($\widehat{\lambda_x^0} < 0$) if the data present positive (negative) skewness, which can be depicted by looking at a data histogram.

Confidence regions for the parameters can be constructed using asymptotic results, assuming that the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$ has approximately a $N_{2p+2}(\boldsymbol{\theta}, \mathbf{J}^{-1}(\boldsymbol{\theta}))$ distribution. In practice, $\mathbf{J}(\boldsymbol{\theta})$ is usually unknown and has to be replaced by the MLE $\mathbf{J}(\widehat{\boldsymbol{\theta}})$. The matrix $\mathbf{J}(\boldsymbol{\theta}) = \sum_{i=1}^n \mathbf{J}_i(\boldsymbol{\theta})$ is required to be positive definite, with $\mathbf{J}_i(\boldsymbol{\theta})$ presented in the Appendix. Our experience in estimating the asymptotic variance of the MLE of λ_x indicates that it typically is overestimated. We believe that more precise confidence intervals are obtained by using the likelihood ratio statistics (see Section 5).

The EM algorithm for the normal model arise as a particular case of the above iterative scheme. All we have to do is to replace $\lambda_x = 0$ in the expressions of the E and M steps and proceed with the simplifications. Moreover, computation of the observed information matrix is accomplished by dropping the third summand in the righthand hand side of (A1). This procedure seems to be important since if λ_x is suspected to be close to zero then the information matrix $\mathbf{J}(\boldsymbol{\theta})$ can be singular, although this has been proved only for simpler models (see, DiCiccio and Monti, 2004).

In simulations conducted, we have noted that when using the *BFGS* method to maximize directly the log-likelihood function (7), many samples the method converge to a point for which the observed information matrix $\mathbf{J}$ is singular. With the EM algorithm this problem does not occur.

4. Hypothesis testing

In the context of the Grubbs model the quality of the measurements is assessed using the additive bias and the precision of the different instruments (the instrument to be preferred is the one with the smallest bias α_j and variance ϕ_j, $j = 1, \ldots, p$). Thus, one hypothesis of interest is to evaluate if the measurements made by the different instruments are exact, that is, $H_{01} : \alpha_2 = \ldots = \alpha_p = 0$. In comparing the precision of the instruments, the hypothesis of interest is $H_{02} : \phi_1 = \ldots = \phi_p$. The hypothesis that consider both situations is $H_{03} : \alpha_2 = \ldots = \alpha_p = 0,\ \phi_1 = \ldots = \phi_p$. Depending on the application, these hypotheses can be tested jointly or separately. Moreover, given rejection in one of these hypothesis, it is frequent to consider testing equality of subsets of variances and bias. We notice that the above hypotheses can be written more generally as

$$H_0 \ : \ \mathbf{C}\boldsymbol{\theta} = \mathbf{d} \tag{19}$$

where $\mathbf{C}$ is ar $r_p \times (2p+2)$ matrix of rank r_C and $\mathbf{d}$ is a r_C-dimensional vector, $\mathbf{C}$ and

$\mathbf{d}$ known. Hypothesis (19) can be tested using the Wald statistic

$$W = (\mathbf{C}\widehat{\boldsymbol{\theta}} - \mathbf{d})^\top (\mathbf{C}^\top \mathbf{J}^{-1}(\widehat{\boldsymbol{\theta}})\mathbf{C})^{-1}(\mathbf{C}\widehat{\boldsymbol{\theta}} - \mathbf{d}). \tag{20}$$

Under (19) and suitable regularity conditions we have that $W \xrightarrow{D} \chi^2_{r_C}$, as $n \to \infty$. Of course, it is also of interest testing the normal model hypothesis, that is, $H_o : \lambda_x = 0$. Although not being formal tests, as in Zhang and Davidian (2001), we compare the SN-GM and the normal model by inspecting some information criteria. Three criteria were selected: the Akaike Information Criterion (AIC, $-\ell(\widehat{\boldsymbol{\theta}}) + P$), Schawarz Bayesian Information Criterion (BIC, $-\ell(\widehat{\boldsymbol{\theta}}) + 0.5\log(N)P$), and the Hannan-Quinn Criterion (HQ, $-\ell(\widehat{\boldsymbol{\theta}}) + \log(\log(N))P$), where P is the number of free parameters in the model and $N = p \times n$. The preferred model is the one with the smallest value of the criterion. Furthermore, we can look at the likelihood ratio statistic (LR) as an informative tool.

5. Application

Now we give an illustrative example of the methodology developed. Barnett (1969) presents data on measurements of the vital capacity of the human lung on a common group of 72 patients, using two instruments (standard and new), operated by skilled and unskilled operators. We consider the measurements divided by 100 in order to improve numerical stability. This data set has been extensively treated in the literature, under using the symmetric structural calibration comparative model of which the Grubbs model is a particular case.

To the EM algorithm, we consider the following convergence criteria

$$\max_{j=1,\ldots,2p+2} \left| \left(\widehat{\theta}_j^{(m+1)} - \widehat{\theta}_j^{(m)} \right) / \widehat{\theta}_j^{(m)} \right| \leq 10^{-4}.$$

Maximum likelihood (ML) estimates and standard errors for the N-GM and SN-GM are given in Table 1. We note that estimates of α_2, α_3 and α_4 and ϕ_j, $j = 1, \ldots, 4$, (and their standard deviations) do not present remarkable differences with both models. Comparing the models by looking at the information criteria, we obtain: AIC=756.8, BIC=773.3 and HQ=755.6 for the SN-GM and AIC=752.9, BIC=771.2 and HQ=751.6 for the normal model. Consequently, the SN-GM outperforms the normal model. The likelihood ratio statistic to test $H : \lambda_x = 0$, is in accordance with the information criteria (LR = 9.78 and P-value = 0.0018). Meanwhile, the Wald statistics (W = 2.20 and P-value=0.13) and the 95% symmetric confidence interval (see Table 1), based on the normal approximation, are in disagreement with the above results. Further insight on differences between the models is provided by considering confidence intervals based on profile likelihoods for each parameter (PCI, Meeker and Escobar, 1995). We note from Table 1 that the PCI for the parameter λ_x seems to provide stronger evidence that is greater that zero and that the PIC for the other parameters are in good agreement with the one based on the normal approximation (CI). This fact is due that the Wald statistic or normal approximation of the ML estimator is only appropriate if the LR is well approximates by a quadratic function (see Pawitan, 2000 and also Figure 2a).

A notable inferential discrepancy, which does not includes the skewness parameter, can be seen in Table 2. When testing the hypothesis $H_{04}: \alpha_2 = \alpha_3 = \alpha_4 = -1.2$, we rejected under SN-GM and not rejected under the N-GM at a 5% level (see Table 2). With the rejection of H_{05} and the acceptation of H_{06} we can say that instrument 2 is better.

Replacing the ML estimates of $\boldsymbol{\theta}$ in (9), we present in Figure (1) Q-Q plots and envelopes (lines represent the 5th percentile, the mean, and the 95th percentile of 100 simulated points for each observation). It seems to us that the plots in Figure 1 provide even stronger evidence than that from the information criteria, that the SN-GM provides a better fit to the data set than the normal Grubbs model.

	N-GM		SN-GM			
Parameter	Estimate	SE	Estimate	SE	CI	PCI
μ_x	22.4611	0.9711	12.1559	1.3918	[9.3720; 14.9396]	[9.46;14.91]
α_2	-0.7042	0.2984	-0.7042	0.2961	[-1.2964; -0.1120]	[-1.29; -0.12]
α_3	-0.9750	0.3610	-0.9750	0.3649	[-1.7048; -0.2452]	[-1.69;-0.26]
α_4	-1.4389	0.3657	-1.4389	0.3693	[-2.1775; -0.7003]	[-2.16;-0.71]
ϕ_x	62.9065	10.6442	168.9100	39.9448	[89.0191; 248.8071]	[129.00;220.00]
ϕ_1	4.9979	0.9947	5.0611	0.9888	[3.0835; 7.0386]	[3.49;7.53]
ϕ_2	1.4129	0.5698	1.2516	0.5640	[0.1235; 2.3797]	[0.29;2.62]
ϕ_3	4.3831	0.9742	4.5264	0.9997	[2.5271; 6.5258]	[2.89;6.95]
ϕ_4	4.6330	1.0321	4.7577	1.0646	[2.6286; 6.8868]	[3.02;7.36]
λ_x	-	-	5.6763	3.8284	[-1.9818; 13.3350]	[2.10; ∞]
log-likelihood	-747.7896		-742.8945			

Tabela 1: Results of fitting N-GM and SN-GM to the data set in Barnett. SE are the estimated asymptotic standard errors. CI is the 95% confidence interval based in the normal approximation of the ML estimates. PCI is the 95% confidence interval based on the profile likelihood.

	N-GM		SN-GM	
Hypothesis	Value	P-value	Value	P-Value
H_{01} : $\alpha_2 = \alpha_3 = \alpha_4 = 0$	16.4007	0.0010	16.1057	0.0011
H_{02} : $\phi_1 = \phi_2 = \phi_3 = \phi_4$	13.9184	0.0009	14.9899	0.0006
H_{03} : $\alpha_2 = \alpha_3 = \alpha_p = 0$, $\phi_1 = \ldots = \phi_4$	30.3191	0.0000	31.0954	0.000
H_{04} : $\lambda_x = 0$	-	-	2.1977	0.1382
H_{05} : $\alpha_2 = \alpha_3 = \alpha_4 = -1.2$	7.6961	**0.0587**	7.9169	**0.0478**
H_{06} : $\alpha_3 = \alpha_4 = -1.2; \phi_1 = \phi_3 = \phi_4$	1.8912	0.7558	1.8022	0.7721

Tabela 2: Barnett data set. Result of the Wald statistic to hypothesis of interest.

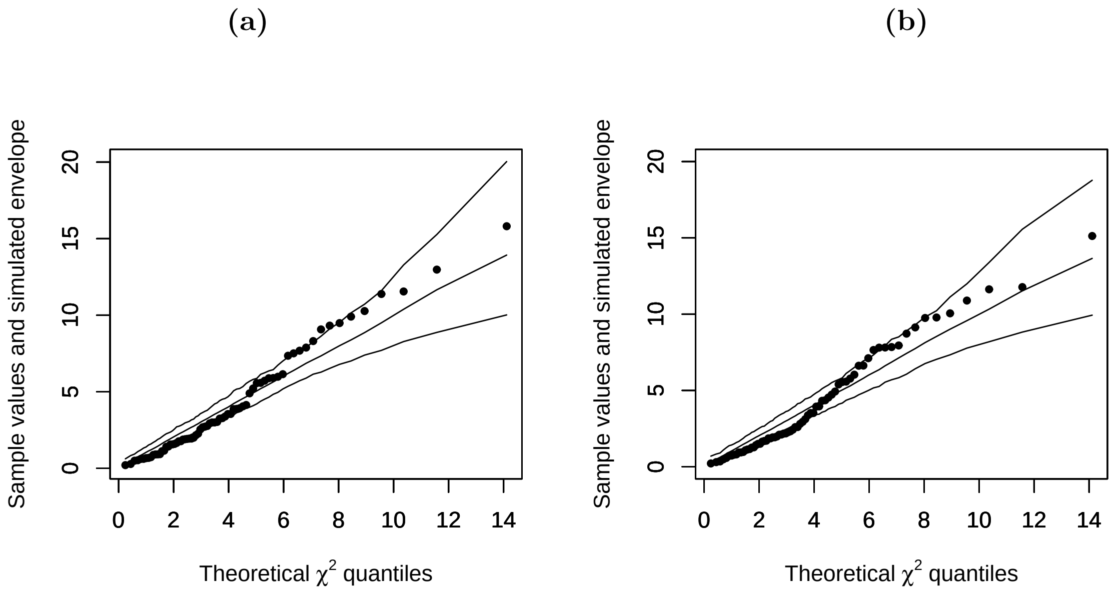

Figura 1: Barnett data set. Q-Q plots and simulated envelopes: (a) Skew-normal Grubbs model and (b) Normal Grubbs model.

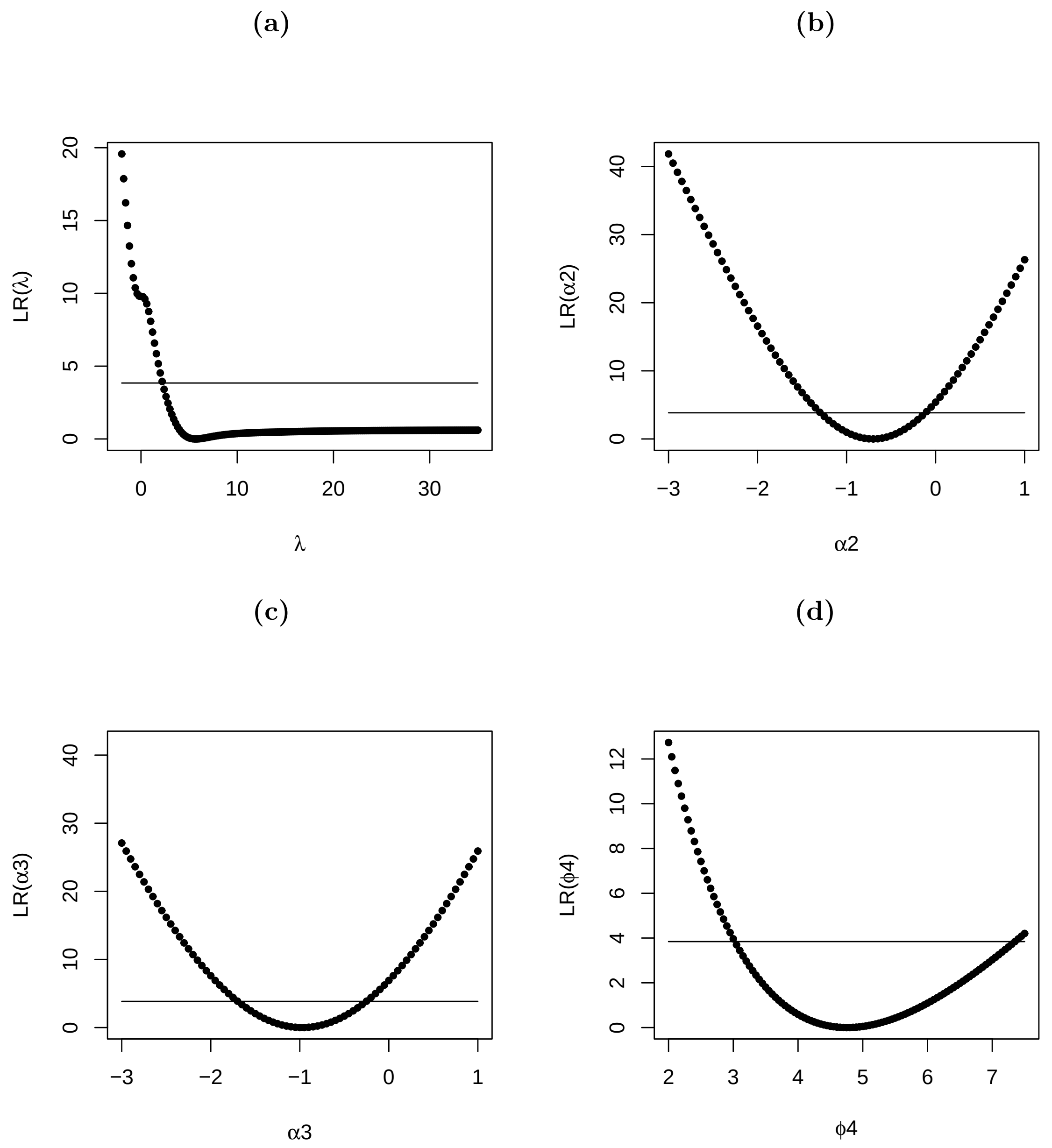

Figura 2: Barnett data set. Likelihood ratio (LR) based on the profile likelihood: (a) λ_x; (b) α_2; (c) α_3 and (d) ϕ_4. The line in each graphic is the 95% limit ($\chi_1^2(0.95) = 3.84$).

6. Conclusions

We have presented strategies to estimation and hypotheses testing in the Grubbs model under the skew-normal distribution. Parameter estimation is conducted via maximum likelihood, by using the EM algorithm, yielding closed form expressions for the equations in the M-step, which seems to be more robust with respect to initial values. Hypothesis testing is approached by using Wald statistics test and we emphasize that it can be less efficient than procedures based on the likelihood. The assessment of influence of data and model assumption on the result of the statistical analysis is a key aspect. Work is in progress addressing specifically influence local and residual analysis.

Acknowledgements

The authors acknowledge the partial financial support from CNPq/CAPES, Brasil.

Appendix

Observed information matrix

In this appendix the observed information matrix is obtained for the SN-GM. Letting $d_i = (\mathbf{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})$ and K_i the last term in (8), we have from (7) that

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}} = -\frac{1}{2}\frac{\partial \mathrm{log}|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\gamma}} - \frac{1}{2} d_i \boldsymbol{\gamma} + \frac{\partial \mathrm{log} K_i}{\partial \boldsymbol{\gamma}}, \tag{A1}$$

where

$$\frac{\partial \mathrm{log} K_i}{\partial \boldsymbol{\gamma}} = W_{\Phi_1}(A_x a_i)\left\{ A_x \frac{\partial a_i}{\partial \boldsymbol{\gamma}} + a_i \frac{\partial A_x}{\partial \boldsymbol{\gamma}} \right\},$$

with $W_{\Phi_1}(u) = \phi_1(u)/\Phi_1(u)$, $u \in \mathbb{R}$, of (7) we can rewrite $K_i = \Phi_1(A_x a_i)$, with $A_x = \dfrac{\phi_x}{c}$ $a_i = (\mathbf{y}_i - \boldsymbol{\mu})^\top D^{-1}(\boldsymbol{\phi})\mathbf{1}_p$ and $c = 1 + \phi_x \mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi})\mathbf{1}_p$, $\boldsymbol{\gamma} = \mu_x, \boldsymbol{\alpha}$, ϕ_x, $\boldsymbol{\phi}$, λ_x, and

$$d_i \boldsymbol{\gamma} = \frac{\partial d_i}{\partial \boldsymbol{\gamma}}, \quad \boldsymbol{\gamma} = \mu_x,\ \boldsymbol{\alpha},\ \phi_x,\ \boldsymbol{\phi},\ \lambda_x,$$

$i = 1, \ldots, n$. Further, using results related to matrix differentiation (Nel, 1980), it follow that

$$\begin{aligned}
\frac{\partial \mathrm{log}|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\gamma}} &= 0, \quad \boldsymbol{\gamma} = \mu_x, \boldsymbol{\alpha}, \lambda_x,\\
\frac{\partial \mathrm{log}|\boldsymbol{\Sigma}|}{\partial \phi_x} &= c^{-1}\frac{c-1}{\phi_x},\\
\frac{\partial \mathrm{log}|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\phi}} &= -\frac{\phi_x}{c} D^{-2}(\boldsymbol{\phi})\mathbf{1}_{\boldsymbol{p}} + D^{-1}(\boldsymbol{\phi})\mathbf{1}_p,\\
\frac{\partial A_x}{\partial \boldsymbol{\gamma}} &= 0, \quad \boldsymbol{\gamma} = \mu_x, \boldsymbol{\alpha},\\
\frac{\partial A_x}{\partial \boldsymbol{\phi}} &= \frac{(2c + \lambda_x^2)}{2\lambda_x^2} A_x^3 D^{-2}(\boldsymbol{\phi})\mathbf{1}_p,\\
\frac{\partial A_x}{\partial \phi_x} &= \frac{(2c + \lambda_x^2 - c^2)}{2\phi_x^2 \lambda_x^2} A_x^3,\\
\frac{\partial A_x}{\partial \lambda_x} &= \frac{\phi_x}{\Lambda_x^2 \lambda_x^3} A_x^3,\\
d_{i\gamma} &= 0, \quad \boldsymbol{\gamma} = \lambda_x,\\
d_{i\mu_x} &= -2\mathbf{1}_{\boldsymbol{p}}{}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i,
\end{aligned}$$

$$
\begin{aligned}
d_{i\boldsymbol{\alpha}} &= -2\mathbb{I}_{(p)}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i,\\
d_{i\phi_x} &= -c^{-2}a_i^2,\\
d_{i\boldsymbol{\phi}} &= -D^{-2}(\boldsymbol{\phi})D(\boldsymbol{X}_i)\boldsymbol{X}_i + 2c^{-1}\phi_x a_i D^{-2}(\boldsymbol{\phi})\boldsymbol{X}_i - c^{-2}\phi_x^2 a_i^2 D^{-2}(\boldsymbol{\phi})\mathbf{1}_p,\\
\frac{\partial a_i}{\partial \boldsymbol{\gamma}} &= 0, \quad \boldsymbol{\gamma} = \phi_x, \lambda_x,\\
\frac{\partial a_i}{\partial \mu_x} &= -\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi})\mathbf{1}_p,\\
\frac{\partial a_i}{\partial \boldsymbol{\alpha}} &= -D^{-1}(\boldsymbol{\psi})\mathbf{1}_{p-1},\\
\frac{\partial a_i}{\partial \boldsymbol{\phi}} &= -D^{-2}(\boldsymbol{\phi})\boldsymbol{X}_i,
\end{aligned}
$$

where $\mathbf{X}_i = (\mathbf{y}_i - \mathbf{a} - \mathbf{1}_p\mu_x)$, $\mathbb{I}_{(p)} = (\mathbf{0}, \mathbf{I}_{p-1})$, of dimension $(p-1)\times p$, and $\boldsymbol{\psi} = (\phi_2, ..., \phi_p)^\top$, $i = 1, \ldots, n$.

From (A1) it follows that the observed, per element, information matrix is given by

$$
\mathbf{J}_i(\boldsymbol{\theta}) = -\frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}^\top} = -\left\{ -\frac{1}{2}\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}^\top} - \frac{1}{2}d_{i\boldsymbol{\gamma}\boldsymbol{\tau}^\top} + \frac{\partial^2 \log K_i}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}^\top} \right\}, \tag{A2}
$$

where

$$
\begin{aligned}
\frac{\partial^2 \log K_i}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}^\top} &= W_{\Phi_1}(A_x a_i)\left\{ \frac{\partial A_x}{\partial\boldsymbol{\gamma}}\frac{\partial a_i}{\partial\boldsymbol{\tau}^\top} + A_x\frac{\partial^2 a_i}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}^\top} + \frac{\partial a_i}{\partial\boldsymbol{\gamma}}\frac{\partial A_x}{\partial\boldsymbol{\tau}^\top} + a_i\frac{\partial^2 A_x}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}^\top} \right\}\\
&\quad + \Delta_{\Phi_1}(A_x a_i)\left\{ A_x\frac{\partial a_i}{\partial\boldsymbol{\gamma}} + a_i\frac{\partial A_x}{\partial\boldsymbol{\gamma}} \right\}\left\{ A_x\frac{\partial a_i}{\partial\boldsymbol{\tau}^\top} + a_i\frac{\partial A_x}{\partial\boldsymbol{\tau}^\top} \right\},
\end{aligned}
$$

$\Delta_{\Phi_1}(u) = W'_{\Phi_1}(u) = -W_{\Phi_1}(u)(u + W_{\Phi_1}(u))$, $u \in \mathbb{R}$, and

$$
d_{i\boldsymbol{\gamma}\boldsymbol{\tau}^\top} = \frac{\partial^2 d_i}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}^\top}, \boldsymbol{\gamma}, \boldsymbol{\tau} = \mu_x, \boldsymbol{\alpha}, \phi_x, \boldsymbol{\phi}, \lambda_x.
$$

After lengthy algebric manipulations we arrive at

$$
\begin{aligned}
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial\boldsymbol{\tau}\partial\boldsymbol{\gamma}^\top} &= 0, \quad \boldsymbol{\tau} = \mu_x, \boldsymbol{\alpha}, \lambda_x; \quad \boldsymbol{\gamma} = \mu_x, \boldsymbol{\alpha}, \phi_x, \boldsymbol{\phi}, \lambda_x,\\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial\phi_x\partial\phi_x} &= -\frac{1}{c^2\phi_x^2}(c-1)^2,\\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial\phi_x\partial\boldsymbol{\phi}^\top} &= -c^{-2}\mathbf{1}_{\boldsymbol{p}}{}^\top D^{-2}(\boldsymbol{\phi}),\\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}^\top} &= -D^{-2}(\boldsymbol{\phi}) - c^{-2}\phi_x^2 D^{-1}(\boldsymbol{\phi})\boldsymbol{M}D^{-1}(\boldsymbol{\phi}) + 2c^{-1}\phi_x D^{-3}(\boldsymbol{\phi}),\\
\frac{\partial^2 A_x}{\partial\phi_x\partial\phi_x} &= -\frac{\lambda_x^2+1}{\lambda_x^2\phi_x^3}A_x^3 + \frac{3(2c+\lambda_x^2-c^2)^2}{4\lambda_x^4\phi_x^4}A_x^5,
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 A_x}{\partial \phi_x \partial \boldsymbol{\phi}^\top} &= [\frac{(c-1)}{\lambda_x^2 \phi_x} A_x^3 + \frac{3(2c+\lambda_x^2)(2c+\lambda_x^2-c^2)}{4\lambda_x^4 \phi_x^2} A_x^5] \mathbf{1_p}^\top D^{-2}(\boldsymbol{\phi}), \\
\frac{\partial^2 A_x}{\partial \phi_x \partial \lambda_x} &= \frac{c-2}{\lambda_x^3 \Lambda_x \phi_x} A_x^3 + \frac{3(2c+\lambda_x^2-c^2)}{2\lambda_x^5 \Lambda_x^2 \phi_x} A_x^5, \\
\frac{\partial^2 A_x}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}^\top} &= [-\frac{\phi_x}{\lambda_x^2} A_x^3 + \frac{3(2c+\lambda_x^2)^2}{4\lambda_x^4} A_x^5] D^{-1}(\boldsymbol{\phi}) \boldsymbol{M} D^{-1}(\boldsymbol{\phi}) \\
&\quad -\frac{2c+\lambda_x^2}{\lambda_x^2} D^{-3}(\boldsymbol{\phi}) A_x^3, \\
\frac{\partial^2 A_x}{\partial \boldsymbol{\phi} \partial \lambda_x} &= \frac{\phi_x A_x^3}{2\lambda_x^5 \Lambda_x^2} [3A_x^2(2c+\lambda_x^2) - 4\lambda_x^2 \Lambda_x] D^{-2}(\boldsymbol{\phi}) \mathbf{1}_p, \\
\frac{\partial^2 A_x}{\partial \lambda_x \partial \lambda_x} &= -\frac{3\phi_x}{\lambda_x^4 \Lambda_x^2} A_x^3 + \frac{3\phi_x^2}{\lambda_x^6 \Lambda_x^4} A_x^5, \\
d_{i\mu_x\mu_x} &= 2\mathbf{1_p}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}_p, \\
d_{i\mu_x \boldsymbol{\alpha}^\top} &= 2\mathbf{1_p}^\top \boldsymbol{\Sigma}^{-1} \mathbb{I}_{(p)}^\top, \\
d_{i\mu_x\phi_x} &= 2\frac{(c-1)}{c^2 \phi_x} a_i, \\
d_{i\mu_x \boldsymbol{\phi}^\top} &= 2c^{-1} \boldsymbol{X}_i^\top \boldsymbol{\Sigma}^{-1} D^{-1}(\boldsymbol{\phi}), \\
d_{i\boldsymbol{\alpha}\boldsymbol{\alpha}^\top} &= 2\mathbb{I}_{(p)} \boldsymbol{\Sigma}^{-1} \mathbb{I}_{(p)}^\top, \\
d_{i\boldsymbol{\alpha}\phi_x} &= 2c^{-2} a_i D^{-1}(\boldsymbol{\psi}) \mathbf{1}_{p-1}, \\
d_{i\boldsymbol{\alpha}\boldsymbol{\phi}^\top} &= 2\mathbb{I}_{(p)} \boldsymbol{\Sigma}^{-1} D^{-1}(\boldsymbol{\phi}) [D(\boldsymbol{X}_i) - c^{-1}\phi_x a_i], \\
d_{i\phi_x\phi_x} &= 2\frac{c^{-3}}{\phi_x}(c-1) a_i^2, \\
d_{i\phi_x \boldsymbol{\phi}^\top} &= (-2c^{-3}\phi_x a_i^2 D^{-2}(\boldsymbol{\phi}) \mathbf{1_p} + 2c^{-2} a_i D^{-2}(\boldsymbol{\phi}) \boldsymbol{X}_i)^\top, \\
d_{i\boldsymbol{\phi}\boldsymbol{\phi}^\top} &= 2D^{-3}(\boldsymbol{\phi}) D^2(\boldsymbol{X}_i) - 4c^{-1}\phi_x a_i D^{-3}(\boldsymbol{\phi}) D(\boldsymbol{X}_i) \\
&\quad -2c^{-1}\phi_x D^{-2}(\boldsymbol{\phi}) \boldsymbol{X}_i \boldsymbol{X}_i^\top D^{-2}(\boldsymbol{\phi}) \\
&\quad +2c^{-2}\phi_x^2 D^{-2}(\boldsymbol{\phi}) \boldsymbol{X}_i \boldsymbol{X}_i^\top \boldsymbol{M} D^{-1}(\boldsymbol{\phi}) \\
&\quad +2c^{-2}\phi_x^2 a_i^2 D^{-3}(\boldsymbol{\phi}) - 2c^{-3}\phi_x^3 a_i^2 D^{-1}(\boldsymbol{\phi}) \boldsymbol{M} D^{-1}(\boldsymbol{\phi}) \\
&\quad +2c^{-2}\phi_x^2 D^{-1}(\boldsymbol{\phi}) \boldsymbol{M} \boldsymbol{X}_i \boldsymbol{X}_i^\top D^{-2}(\boldsymbol{\phi}), \\
\frac{\partial^2 a_i}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} &= 0, \quad \boldsymbol{\gamma} = \boldsymbol{\phi}, \quad \boldsymbol{\tau} = \phi_x, \lambda_x, \\
\frac{\partial^2 a_i}{\partial \mu_x \partial \boldsymbol{\phi}^\top} &= \mathbf{1_p}^\top D^{-2}(\boldsymbol{\phi}), \\
\frac{\partial^2 a_i}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\phi}^\top} &= \mathbb{I}_{(p)} D^{-2}(\boldsymbol{\phi}), \\
\frac{\partial^2 a_i}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}^\top} &= 2D(\boldsymbol{X}_i) D^{-3}(\boldsymbol{\phi}).
\end{aligned}
$$

$i = 1, \ldots, n$, where $\boldsymbol{M} = D^{-1}(\boldsymbol{\phi}) \mathbf{1}_p \mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi})$.

References

Arellano-Valle, R. B. and Genton, M. G. (2005). Fundamental skew distributions. *Journal of Multivariate Analysis,* **96**, 93-116.

Arellano-Valle R. B., Bolfarine, H. and Lachos, V. H. (2005a). Skew-normal linear mixed models. *Journal of Data Science,* **3**, 415-438.

Arellano-Valle R. B., Ozan, S., Bolfarine, H. and Lachos, V. H. (2005b). Skew-normal measurement error models. *Journal of Multivariate Analysis,* **96**, 265-281.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics,* **12**, 171-178.

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica (Bologna),* **46**, 199-208.

Azzalini, A. and Capitanio, A. (2005). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society - Series B,* **61**, 579-602.

Azzalini, A. and Dalla-Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika,* **83**, 715-726.

Barnett, V. D. (1969). Simultaneous pairwise linear structural relationships. *Biometrics,* **25**, 129-142.

Bazán, J. L., Branco, M. D. and Bolfarine, H. (2005). A skew item response model. To appear in *Bayesian Analysis.*

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society - Series B,* **39**, 1-22.

Dalla Valle, A. (2004). The skew-normal distribution. In: Genton, M. G. *Skew-Elliptical Distributions and their Applications,* (Boca Raton, FL: Chapman & Hall/CRC), pp 3-24.

DiCiccio, T. J. and Monti, A. C. (2004). Inferential aspects of the skew exponential power distribution. *Journal of the American Statistical Association,* **99**, 439-450.

Genton, M. G. (2004). *Skew-Elliptical Distributions and their Applications,* Boca Raton, FL: Chapman & Hall/CRC.

Grubbs, F. E. (1948). On estimating precision of measuring instruments and product variability. *Journal of the American Statistical Association,* **43**, 243-264.

Grubbs, F. E. (1973). Errors of measurement, precision, accuracy and the statistical comparison of measuring instruments. *Technometrics,* **15**, 53-66.

Grubbs, F. E. (1983). Grubbs estimator. *Encyclopedia of Statistical Sciences,* **3**, 542-549.

Henze, N. (1986). A probabilistic representation of the skew-normal distribution. *Scandinavian Journal of Statistics,* **13**, 271-275.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions, volume 1,* New York, NJ: Wiley.

Lachos, V. H, Bolfarine, H., Vilca, F. and Galea (2005). Estimation and influence diagnostic for structural comparative calibration models under the skew-normal distribution. Technical Report RT-MAE, 2005-11, IME-USP, São Paulo, Brazil.

Meeker, W.Q. and Escobar, L. A. (1995). "Teaching about approximate confidence regions based on maximum likelihood estimation". *The American Statistician,* **49**, 48-53.

Nel, D. G. (1980). On matrix differentiation in statistics. *South African Statistical Journal,* **14**, 137-193.

Pawitan, Y. (2000). A reminder of the fallibility of the Wald statistic: likelihood explanation. *The American Statistician,* **54**, 54-56.

Zhang, D. and Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics,* **57**, 795-802.

Vilca, F., Lachos, V.H. and Bolfarine, H.(2002). On testing statistics for comparing several measuring devices. Technical Report RP-IMECC, 2002-51, IMECC-UNICAMP, São Paulo, Brazil.